METHOD FOR PREPARATION OF CDNA TAGS FOR IDENTIFYING EXPRESSED GENES AND METHOD FOR ANALYSIS OF GENE EXPRESSION

Cross-Reference To Related Application

This application claims benefit of and priority to Japanese Patent Application No. 2002-267163 filed on September 12, 2002, and which is incorporated by reference in its entirety.

Technical Field

The present invention relates to a method for the preparation of cDNA tags for identifying expressed genes, a cDNA library prepared by the method and a method for the analysis of gene expression. More specifically, the invention relates to the method for the preparation of cDNA tags hybridizing to mRNAs as products of expressed genes, cDNAs corresponding to the mRNAs or given areas of the cDNA fragments, and the method for analysis of gene expression using the cDNAs.The method for the analysis of gene expression includes a direct method using the cDNA tags without any processing and an indirect method using a concatemer of the cDNA tags.

Background Art

Each species has the peculiar gene expression pattern based on the original genomic sequence. In addition, even if the species is the same, it has been found that each cell or organ shows different gene expression patterns

depending on a physiological stage such as degree of differentiation, multiplication and aging, or a pathological state such as canceration, infectious disease and immunologic disease. Accordingly, when such gene expression patterns are compared with each other, the differences provide valuable information which can be used in a variety of applications such as identification of appropriate treatment targets, identification of candidate genes for a gene therapy, tissue typing, legal gene confirmation, positioning of disease related genes and identification of indicator genes for diagnosis.

Northern blotting method, RNase protection method and Reverse transcriptase-polymerase chain reaction (RT-PCR) analysis method (Alwine et al., Proc. Natl. Acad. Sci. U.S.A., 74:5350, 1977; Zinn et al., Cell, 34:865, 1983; Veres et al., Science, 237:415, 1987) were designed in order to evaluate gene expressions. Further methods good for retrieving genes such as expressed sequence tag: EST (Adams et. al, Science 252:1651, 1991; Adams et. al., Nature,355: 632, 1992; Okubo et.al., Nature Genetics, 2:173, 1992) have been developed. However the methods can evaluate the limited number of genes at a time. For example, Okubo et al. developed a method for obtaining a profile of gene expression comprising the steps of cleaving double-stranded cDNAs with the restriction enzyme having a four base recognition site (MboI) to prepare a cDNA library consisting of 3'-end fragments of the mRNAs, cloning the 3'-end

fragments and then sequencing randomly (Okubo et al., Nature Genetics, 2: 173, 1992). Since the method provides the clones having the length of about 300 bases on average and requires sequencing each of the clones separately, a total number of the mRNAs that were finally sequenced in a cell were only about 1,000. As a result, the profile from the method was far from the true pattern of the gene expression. Further, since these methods need lots of samples (for example human tissue), cause bias of the results by repeating the polymerase chain reaction (PCR) and lack reproducibility of the results, they have been used merely in a laboratory.

Recently, a method for the serial analysis of gene expression (referred to as SAGE) has been developed (see WO97/10363, US Patents Nos. 5,695,937 and 5,866,330). The SAGE can analyze lots of transcripts by identifying a given region of the transcripts corresponding to the expressed genes. In this method, the patterns of gene expression are determined by preparing tags referred to as "ditag" which are formed by ligating randomly two of short nucleotides corresponding to each cDNA in a sample, connecting the ditags like a chain to form concatemers, cloning and sequencing each concatemer to determine the pattern of gene expression. The SAGE cannot provide a single cDNA tag for identifying expressed gene corresponding to each cDNA in a sample and the number of expressed genes to be identified at a time is limited to 1,000 or less,

usually 400 or less because the concatemer can contain the limited number of the ditags.

In order to solve the drawbacks of conventional methods for the analysis of gene expression patterns and to improve the low analysis capability of the methods, the inventors have developed a method for the preparation of cDNA tags for identifying expressed genes and a method for the analysis of gene expression using the cDNA tags, which are described in PCT/JP02/02338.

Disclosure of Invention

International patent application No.PCT/JP02/02338 which was filed by the same inventors of the present application discloses a method for the preparation of cDNA tags for identifying expressed genes and a method for the analysis of gene expression using the cDNA tags.

An object of the present invention is to provide an improvement of the method of PCT/JP02/02338 in that, for example the restriction enzymes used therein can be selected from a broader aspect than those of the previous invention and the linkers used therein can be prepared more easily as well as provide the better method for the analysis of gene expression.

The present invention provides a method for the preparation of cDNA tags for identifying expressed genes, which enables one to conduct the efficient analysis of peculiar gene expression pattern of each species, and of specific gene

expression patterns depending on a physiological state, on a development step, or on a pathological states of cells or organs, and also provides a method for the gene expression using the cDNA tags. The method of the present invention requests a less amount of cell samples for the analysis of gene expression and is more efficient and reliable than the conventional technologies.

The present invention provides a method for the preparation of cDNA tags for identifying expressed genes comprising: providing complementary deoxyribonucleic acids (cDNAs); cleaving the cDNAs with a type II restriction enzyme to prepare cDNA fragments; ligating the cDNA fragments to linker Xes which have a recognition site of a first type IIS restriction enzyme and which form a recognition site of a second type IIS restriction enzyme at the site linking with the cleavage end of the cDNA fragments formed by the type II restriction enzyme to prepare linker X-cDNA fragment complexes; cleaving the linker X-cDNA fragment complexes with the second type II restriction enzyme to prepare linker X-cDNA tag complexes; ligating linker Ys which have a recognition site of a third type IIS restriction enzyme to the cleavage end of the linker X-cDNA tag complexes formed by the second type IIS restriction enzyme to prepare linker X-cDNA tag-linker Y complexes; amplifying the linker X-cDNA fragment-linker Y complexes; and cleaving the amplified products thus obtained with the first and third type IIS restriction enzymes simultaneously or in turn to prepare the cDNA

tags for identifying expressed genes.

The present invention also provides a method for the preparation of cDNA tags for identifying expressed genes comprising: providing complementary deoxyribonucleic acids (cDNAs); cleaving the cDNAs with a type II restriction enzyme to produce cDNA frangments; ligating the cDNA fragments to linker Xes which have recognition sites of first and second type IIS restriction enzymes to prepare linker X-cDNA fragment complexes; cleaving the linker X-cDNA fragment complexes with the second type IIS restriction enzyme to prepare linker X-cDNA tag complexes; ligating linker Ys which have a recognition site of the first type IIS restriction enzyme to the cleavage end of the linker X-cDNA tag complexes formed by the second type IIS restriction enzyme to prepare linker X-cDNA tag-linker Y complexes; amplifying the linker X-cDNA tag-linker Y complexes; and cleaving the amplified products thus obtained with the first IIS restriction enzyme to prepare the cDNA tags for identifying expressed genes.

The present invention further provides a method for the preparation of cDNA tags for identifying expressed genes comprising: providing complementary deoxyribonucleic acids; cleaving the cDNAs with a type II restriction enzyme to produce cDNA fragments; ligating the cDNA fragments to linker Xes which have recognition sites of first and second type IIS restriction enzymes to prepare linker X-cDNA fragment complexes; cleaving the linker X-cDNA fragment complexes

with the second type IIS restriction enzyme to prepare linker X-cDNA tag complexes; ligating linker Ys which have a recognition site of a third type IIS restriction enzyme to the cleavage end of the linker X-cDNA tag complexes formed by the second type IIS restriction enzyme to prepare linker X-cDNA tag-linker Y complexes; amplifying the linker X-cDNA tag-linker Y complexes; and cleaving the amplified products thus obtained with the first and third type IIS restriction enzymes simultaneously or in turn to prepare the cDNA tags for identifying expressed genes.

The present invention further provides linker X comprising recognition sites of the first and second type IIS restriction enzymes.

The present invention further provides a method for the analysis of gene expression wherein the library of cDNA tags prepared by the method described above is contacted with a detector on which nucleic acids to be detected are immobilized.

The present invention further provides a method for the analysis of gene expression wherein a library of nucleic acids to be detected is connected with the detector on which the cDNA tags prepared by the method of the present invention are immobilized. DNA chips may be used as the detector therefore.

The present invention further provides a method for the analysis of gene expression comprising the steps of concatenating the cDNA tags for identifying

expressed genes prepared by the method described above to form concatemers and sequencing the concatemers. The method for the analysis includes a method for the qualitative analysis of gene expression wherein the concatemers are sequenced and then each of the cDNA tags are sequenced on the basis of the sequences of the concatemers, and a method for the quantitative analysis of gene expression wherein each of the cDNA tags are sequenced and measured in frequency of occurrences on the basis of the sequences of the concatemers.

The present invention further provides a kit for the preparation of cDNA tags for identifying expressed genes wherein the kit comprises a type II restriction enzyme, a first type IIS restriction enzyme, a second type IIS restriction enzyme, a third type IIS restriction enzyme, linker Xes which have a recognition site of the first type IIS restriction enzyme and which form a recognition site of the second type IIS restriction enzyme at the site linking with the cleavage end of the cDNA fragments formed by the type II restriction enzyme to prepare linker X-cDNA fragment complexes and linker Ys which have a recognition site of the third type IIS restriction enzyme.

The present invention further provides a kit for the preparation of cDNA tags for identifying expressed genes wherein the kit comprises a type II restriction enzyme, a first type IIS restriction enzyme, a second type IIS restriction enzyme, linker Xes which have recognition sites of the first and the

second type IIS restriction enzymes and linker Ys which have a recognition site of the first type IIS restriction enzyme.

In addition, the present invention further provides a kit for the preparation of cDNA tags for identifying expressed genes wherein the kit comprises a type II restriction enzyme, a first type IIS restriction enzyme a second type IIS restriction enzyme, a third type IIS restriction enzyme, linker Xes which have recognition sites of the first and second type IIS restriction enzymes and linker Ys which have a recognition site of the third type IIS restriction enzyme.

The term "cDNA tag for identifying expressed gene" as used herein may be abbreviated as cDNA tag for IEG, if necessary.

Fundamental principles of the present invention will be explained hereinafter. First, a short nucleotide sequence isolated from a defined region within a gene transcript has sufficient information to identify the transcript. For example, a sequence of 9 bp may have combinations of the ninth power of four, 262,144 and therefore the sequence can identify the same number of the transcripts. Whereas, estimates suggest that the human genome encodes about 80,000 to 200,000 transcripts (Fields, et al Nature Genetics, 1:345 1994). Principally, if the tags of 9bp are obtained, all of the transcripts of the human genome can be identified.

The size of the tag may be shorter, where a subject of the analysis is a

lower eukaryote or prokaryote, because the number of transcripts encoded by the genome is lower. For example, a tag of 6 to 7 bp may be sufficient for distinguishing the transcripts in yeast. The present invention can provide single cDNA tags of the same length for identifying expressed genes with a variety of lengths and therefore is useful in the analysis of gene expression patterns.

Second, the present invention can provide extremely reduced bias caused by amplification and/or cloning because the invention allows analyzing a gene expression by once amplification of a single short cDNA tag interposed between upstream and downstream linkers.

Third, a library of the cDNA tags prepared according to the present invention can be used to qualitatively or quantitatively detect the cDNAs corresponding to the cDNA tags for the analysis of gene expression patterns.

Fourth, concatemers with or without spacer sequences of the cDNA tags prepared by the method of the present invention allows serial and efficient analysis of gene expression. If necessary, the concatemers may be cloned into a vector and the like. Specifically, since the cDNA tags have independent sequences individually, it is easy to sequence each of the concatemers and to singly isolate the cDNA tags from the concatemers.

It is common between the present invention and the SAGE method described hereinbefore in the first principle that a tag of short nucleotide

sequence has sufficient information to identify the transcript. However the SAGE uses a complexized tag referred to as "ditag". The SAGE is different from the present invention in that the SAGE does not prepare and use a single cDNA tag for identifying expressed gene of the present invention, a library thereof and a concatemer of the single cDNA tags.

These and other objects, advantages, and features of the invention will become apparent to those skilled in the art upon reading the details of the invention as more fully described below.

Brief Description of Drawings

Each figure is a process drawing, which shows the process of embodiment of the method for the preparation of cDNA tags for identifying expressed genes according to the present invention. The letter "N" in the figures means any one of A, T, G, or C.

Figure 1 shows steps (1) to (6) of the method for the preparation of cDNA tags according to the invention using linker X having a recognition site of third type IIS restriction enzyme.

Figure 2 shows steps (7) to (10) of the method for the preparation of cDNA tags according to the invention using linker Y having a recognition site of a third type IIS restriction enzyme.

Figure 3 shows steps (1) to (6) of the method for the preparation of cDNA

tags according to the invention using linker X having recognition sites of first and second type IIS restriction enzymes.

Figure 4 shows steps (7) to (10) of the method for the preparation of the cDNA tags according to the invention using linker X having recognition sites of first and second type IIS restriction enzymes.

Figure 5 shows steps (1) to (6) of the method for the preparation of the cDNA tags according to the invention using linker X having recognition sites of first and second type IIS restriction enzymes and linker Y having a recognition site of third type IIS restriction enzyme.

Figure 6 shows steps (7) to (10) of the method for the preparation of the cDNA tags according to the invention using linker X having recognition sites of first and second type IIS restriction enzymes and linker Y having a recognition site of third type IIS restriction enzyme.

Modes for Carrying Out the Invention

According to a preferred embodiment of the present invention, the method for the preparation of cDNA tags for identifying expressed genes, which is referred to as "cDNA tag" or "cDNA tag for IEG" hereinbelow, will be explained in detail using flow-charts shown in figures 1 to 6. In this connection, the method for the preparation of cDNA tags according to the present invention has three modes.

That is, the present invention provides the following modes: a method comprising the step of ligating linker Ys which have a recognition site of a third type IIS restriction enzyme to the cleavage end of the linker X-cDNA tag complexes formed by a second type IIS restriction enzyme (mode 1); a method comprising the step of ligating linker Xes which have recognition sites of first and second type IIS restriction enzymes to given cDNA fragments (mode 2); and a method comprising a combination of the both steps using said linker Xes and linker Ys (mode 3).

The method of the preparation of the cDNA tags according to mode 1 of the present invention are illustrated below on the basis of figures 1 and 2.

The method comprising the steps of:

(1) providing complementary deoxyribonucleic acids (cDNAs);

(2) cleaving the cDNAs with a type II restriction enzyme RsaI to prepare cDNA fragments;

(3) ligating the cDNA fragments to linker Xes which have a recognition site of a first type IIS restriction enzyme BseRI and which form a recognition site of a second type IIS restriction enzyme BsmFI at the site linking with the cleavage end of the cDNA fragments formed by the type II restriction enzyme RsaI to prepare linker X-cDNA fragment complexes;

(4) cleaving the linker X-cDNA fragment complexes with the second type

IIS restriction enzyme BsmFI to prepare linker X-cDNA tag complexes;

(5) refining the linker X-cDNA tag complexes, if necessary;

(6) processing the cleavage end of the cDNA tags of the linker X-cDNA tag complexes to make the end capable of binding the linker Ys which have a recognition site of the third type IIS restriction enzyme EciI;

(7) ligating linker Ys which have the recognition site of the third type IIS restriction enzyme EciI to the cleavage end of the linker X-cDNA tag complexes formed by the second type IIS restriction enzyme BsmFI to prepare linker X-cDNA tag-linker Y complexes;

(8) amplifying the linker X-cDNA fragment-linker Y complexes;

(9) cleaving the amplified products thus obtained with the first type IIS restriction enzymes BseRI and the third type IIS restriction enzyme EciI simultaneously or in turn to prepare the cDNA tags for identifying expressed genes; and

(10) isolating the obtained cDNA tags, if necessary.

Although the concrete restriction enzymes are mentioned in the description of mode 1 to make the invention more understandable, of course, mode 1 of the invention can use many kinds of restriction enzymes. The term "linker X" as used in mode 1 means a linker which has a recognition site of a first type IIS restriction enzyme and which forms a recognition site of a second

type IIS restriction enzyme at the site linking with the cleavage end of the cDNA fragments formed by the type II restriction enzyme. The term "linker Y" means a linker having a recognition site of a third type IIS restriction enzyme.

Specific sequences of linker X, linker Y, primer X and primer Y used in mode 1 are shown below.

Linker X:

5'-···NNNNNNNGAGGAGNNNNNGGG···-3' (SEQ ID NO: 1)

3'-···NNNNNNNCTCCTCNNNNNCCC···-5' (SEQ ID NO: 2)

Linker Y:

5'-··· CNNNNNNNNNTCCGCCNNNNNNN···-3' (SEQ ID NO: 3)

3'-···TGNNNNNNNNNAGGCGGNNNNNNN···-5' (SEQ ID NO: 4)

Primer X:

5'-···NNNNNNNGAGGAGNNNNNGGGAC···-3' (SEQ ID NO: 5)

Primer Y:

3'-···TGNNNNNNNNNAGGCGGNNNNNNN···-5' (SEQ ID NO: 6)

The method for the preparation of cDNA tags according to mode 2 of the present invention are illustrated below on the basis of Figures 3 and 4. The method comprising the steps of:

(1) providing complementary deoxyribonucleic acids (cDNAs);

(2) cleaving the cDNAs with a type II restriction enzyme Csp6I to prepare cDNA fragments;

(3) ligating the cDNA fragments to linker Xes which have recognition sites of first type IIS restriction enzyme BseRI and second type IIS restriction enzyme BsgI to prepare linker X-cDNA fragment complexes;

(4) cleaving the linker X-cDNA fragment complexes with the second type IIS restriction enzyme BsgI to prepare linker X-cDNA tag complexes;

(5) refining the linker X-cDNA tag complexes, if necessary;

(6) processing the end of cDNA tags of the linker X-cDNA tag complexes to make the end capable of binding the linker Ys which have a recognition site of third type IIS restriction enzyme;

(7) ligating linker Ys which have a recognition site of second type IIS restriction enzyme BseRI to the cleavage end of the linker X-cDNA tag complexes formed by second type IIS restriction enzyme BsgI to prepare the linker X-cDNA tag-linker Y complexes;

(8) amplifying the linker X-cDNA tag-linker Y complexes;

(9) cleaving the amplified products thus obtained with the first type IIS restriction enzyme BseRI to prepare the cDNA tags; and

(10) isolating the obtained cDNA tags, if necessary.

Although the concrete restriction enzymes are also mentioned in the

description of mode 2 to make the invention more understandable, mode 2 of the invention can use many kinds of restriction enzymes. The term "linker X" as used in the description of mode 2 means a linker which has a recognition site of first type IIS restriction enzyme and a recognition site of second type IIS restriction enzyme BsgI. The term "linker Y" means a linker having the same recognition site of the first type IIS restriction enzyme used in the linker X.

Specific sequences of linker X, linker Y, primer X and primer Y used in mode 2 are shown below.

Linker X:

5'-···NNNNNNNGAGGAGNGTGCAG···-3' (SEQ ID NO: 7)

3'-···NNNNNNNCTCCTCNCACGTCAT···-5' (SEQ ID NO: 8)

Linker Y:

5'-··· ACNNNNNNNNCTCCTCNNNNNNN···-3' (SEQ ID NO: 9)

3'-···NNTGNNNNNNNNNGAGGAGNNNNNNN···-5' (SEQ ID NO: 10)

Primer X:

5'-···NNNNNNNGAGGAGNGTGCAGTAC···-3' (SEQ ID NO: 11)

Primer Y:

3'-···TGNNNNNNNNGAGGAGNNNNNNN···-5' (SEQ ID NO: 12)

The method for the preparation of the cDNA tags according to mode 3 of

the present invention are illustrated below on the basis of figures 5 and 6. The method comprising the steps of:

(1) providing complementary deoxyribonucleic acids (cDNAs);

(2) cleaving the cDNAs with a type II restriction enzyme Csp6I to prepare cDNA fragments;

(3) ligating the cDNA fragments to linker Xes which have recognition sites of first type IIS restriction enzyme BseRI and second type IIS restriction enzyme BsgI to prepare linker X-cDNA fragment complexes;

(4) cleaving the linker X-cDNA fragment complexes with second type IIS restriction enzyme BsgI to prepare linker X-cDNA tag complexes;

(5) refining the linker X-cDNA tag complexes, if necessary;

(6) processing the cleavage end of cDNA tags of the linker X-cDNA tag complexes to make the end-sites capable of binding the linker Ys which have a recognition site of third type IIS restriction enzyme EciI;

(7) ligating linker Ys which have a recognition site of third type IIS restriction enzyme Ecil to the cleavage end of the linker X-cDNA tag complexes formed by second type IIS restriction enzyme, BsgI to prepare linker X-cDNA tag-linker Y complexes;

(8) amplifying the linker X-cDNA fragment-linker Y complexes;

(9) cleaving the amplified products thus obtained with first type IIS

restriction enzymes BseRI and third type IIS restriction enzyme EciI simultaneously or in turn to prepare the cDNA tags; and

(10) isolating the obtained cDNA tags, if necessary.

Although the concrete restriction enzymes are also mentioned in the description of mode 3 to make the invention more understandable, of course, mode 3 of the invention can utilyze many kinds of restriction enzymes. The term "linker X" as used in the description of mode 3 means a linker which has a recognition site of first type IIS restriction enzyme and a recognition site of second type IIS restriction enzyme. The term "linker Y" means a linker having a recognition site of third type IIS restriction enzyme.

The present invention will be explained more specifically according to mode 3. The restriction enzymes and procedures described below may be applicable to the entire invention, unless otherwise indicated.

In step (1), cDNAs are prepared as a sample. Normally, mRNAs are prepared from the cells to be examined and then the cDNAs are prepared by conventional procedure with a reverse transcriptase. The cDNAs may be sequences corresponding to the full-length mRNAs, fragments of the mRNAs or a combination thereof. The cells to be examined are not limited and may be all the cells including animal cells, plant cells, and microbial cells, where the cells produce mRNAs having a poly A tail at the 3' end. Virus-infected animal cells,

plant cells or microbial cells may also be used as the cells to be examined in the invention.

A method of the present invention can analyze a gene expression when one microgram (ug) of mRNAs from the sample is available. Since 1 ug of mRNAs can be normally obtained from 1 mg of the cells to be examined, the present invention is particularly useful in handling precious human tissue samples obtained by needle biopsy.

Isolation of mRNAs from the cells to be examined may be performed by conventional procedure. For example, the mRNAs are obtained by treating the cells with a guanidine reagent or a phenol reagent to isolate total RNAs and then performing an affinity-column method or a batch method using an oligo dT-cellulose or Sepharose 2B as a carrier.

The first chain cDNAs (single-stranded cDNAs) are synthesized with the resultant mRNAs as a template using oligo dT primers and a reverse transcriptase and then the second chain cDNAs (double-staranded cDNAs) are synthesized with the first chain cDNAs as a template.

The oligo dT primers used herein include an oligo dT primer immobilized on a solid phase and an oligo dT primer labeled with a coenzyme marker. In light of reproducibility and recovery rate of the targeted cDNA fragments, the oligo dT primer immobilized on a solid phase is preferable. The oligo dT primers

immobilized on a solid phase may be oligo dT primers immobilized on latex beads or oligo dT primers immobilized on magnet beads, preferably the oligo dT primers immobilized on magnet beads.

In step (2), the cDNAs in the sample are cleaved with a type II restriction enzyme to prepare the cDNA fragments. The cDNAs in the sample may be double-stranded cDNAs combined with the oligo dT primers immobilized on a solid phase. The term "type II restriction enzyme" as used herein means a restriction enzyme which recognizes a given recognition site and then cleaves the DNA at a specific position inside of or adjacent to the recognition site. The type II restriction enzyme used in the present invention is a restriction enzyme having at least one recognition site of the mRNA to be analyzed, for example, preferably a type II restriction enzyme having a recognition site consisting of 4, 5 or 6 bases.

In particular, taking an average length of the mRNAs, about 2,000 bases into consideration, a type II restriction enzyme having a recognition site of four bases is preferable to the invention because the enzyme theoretically has a recognition site per the fourth power of four, 256 bases.

The type II restriction enzymes used in the invention include AfaI, AluI, CfuI, CviRI, DpnI, EsaBC3I, HpyBI, HpyCH4V, HpyF44III, MltI, PlaAII, RsaI, BfaI, Csp6I, CviAII, CviQI, CviRII, FgoI, HpyCH4IV, MaeI, MaeII, MthZI, RmaI, PpaAII, TaqI, Tsp32I, Tsp32II, TthHB8I, XspI, BspKT6I, BstKTI, HpyCH4I,

AspMDI, Bce243I, Bfi57I, BfuCI, Bme12I, BscFI, Bsp67I, Bsp105I, Bsp143I, Bsp2095I, BspAI, BspFI, BspJI, Bst19II, BstENII, BtkII, CacI, CcyI, ChaI, CpfI, CviAI, DpnII, FatI, FnuCI, FnuEI, HacI, Kzo9I, LlaAI, MboI, MgoI, MkrAI, NdeII, NlaII, NmeCI, NphI, RalF40I, Sau3AI, SauMI, Sth368I, ChaI, Hn1II, Hsp92II, NlaIII, TaiI, TscI, Tsp49I, AccII, BanAI, BceBI, BecAII, BepI, Bim19II, Bme361I, Bpu95I, BseQI, BshI, Bsh1236I, BshFI, Bsp50I, Bsp123I, Bsp211I, BspBRI, BspKI, BspRI, BstFNI, BstUI, Bsu1532I, BsuRI, BtgI, BtkI, CltI, Csp68KVI, CspKVI, DsaII, EsaBC4I, FalII, FauBII, FnuDI, FnuDII, HaeIII, MchAII, MfoAI, MfoAI, MvnI, NgoPII, NspLKI, PalI, Pde133I, PflKI, PlaI, SbvI, SfaI, SuaI, ThaI, AciI, Bco27I, BsiSI, Bst40I, BsuFI, CboI, HapII, Hin2I, Hin6I, HinP1I, HpaII, HsoI, HspAI, MnoI, MspI, Pde137I, SciNI, Sth134I, AspLEI, BspLAI, BstHHI, CfoI, FnuDIII, HhaI, SelI, MseI, Tru1I, Tru9I, Sse9I, TasI, Tsp509I and TspEI.

These type II restriction enzymes have a recognition site consisting of all the four bases, A, T, C and G or other recognition site consisting of C and G or A and T.

The type II restriction enzyme having a recognition site consisting of all the four bases, A, T, C and G include AfaI, AluI, CfuI, CviRI, DpnI, EsaBC3I, HpyBI, HpyCH4V, HpyF44III, MltI, PlaAII, RsaI, BfaI, Csp6I, CviAII, CviQI, CviRII, FgoI, HpyCH4IV, MaeI, MaeII, MthZI, RmaI, PpaAII, Tsp32I, Tsp32II,

TaqI, TthHB8I, XspI, BspKT6I, BstKTI, HpyCH4I, AspMDI, Bce243I, Bfi57I, BfuCI, Bme12I, BscFI, Bsp67I, Bsp105I, Bsp143I, Bsp2095I, BspAI, BspFI, BspJI, Bst19II, BstENII, BtkII, CacI, CcyI, ChaI, CpfI, CviAI, DpnII, FatI, FnuCI, FnuEI, HacI, Kzo9I, LlaAI, MboI, MgoI, MkrAI, NdeII, NlaII, NmeCI, NphI, RalF40I, Sau3AI, SauMI, Sth368I, ChaI, Hn1II, Hsp92II, NlaIII, TaiI, TscI and Tsp49I.

The type II restriction enzymes having a recognition site consisting of bases C and G include AccII, BanAI, BceBI, BecAII, BepI, Bim19II, Bme361I, Bpu95I, BseQI, BshI, Bsh1236I, BshFI, Bsp50I, Bsp123I, Bsp211I, BspBRI, BspKI, BspRI, BstFNI, BstUI, Bsu1532I, BsuRI, BtgI, BtkI, CltI, Csp68KVI, CspKVI, DsaII, EsaBC4I, FalII, FauBII, FnuDI, FnuDII, HaeIII, MchAII, MfoAI, MfoAI, MvnI, NgoPII, NspLKI, PalI, Pde133I, PflKI, PlaI, SbvI, SfaI, SuaI, ThaI, AciI, Bco27I, BsiSI, Bst40I, BsuFI, CboI, HapII, Hin2I, Hin6I, HinP1I, HpaII, HsoI, HspAI, MnoI, MspI, Pde137I, SciNI, Sth134I, AspLEI, BspLAI, BstHHI, CfoI, FnuDIII, HhaI and SelI.

In addition, the type II restriction enzymes having a recognition site consisting of bases A and T include MseI, Tru1I, Tru9I, Sse9I, TasI, Tsp509I and TspEI. The restriction enzyme is preferably selected in light of the features of these recognition sites.

In step (3), linker Xes having recognition sites of the first and second type

IIS restriction enzymes are ligated to the cDNA fragment to prepare the linker X-cDNA fragment complexes. cDNA fragments having an oligo dT primer sequence are isolated from the group of the cDNA fragments prepared in step (2). The isolation may be performed by using a label of oligo dT primer. For example, where oligo dT primers immobilized on latex beads are used for the preparation of said cDNAs, the cDNAs may be treated with type II restriction enzyme, centrifuged to precipitate in the form of cDNA fragments having the oligo dT primer sequences immobilized on the beads and then isolated. The cDNA fragments thus obtained are those having a poly A tail and a cleavage-end of said type II restriction enzyme which first appears in the 5' upstream direction from the poly A tail. In the next process, the cDNA fragments are ligated to linker Xes using DNA ligase such as T4 DNA ligase.

A recognition site of the first type IIS restriction enzyme in linker Xes may be located so that the first type IIS restriction enzyme cleaves the cDNA tag leaving no spacer sequence or a desired spacer sequence at an appropriate position.

For example, where linker X has a recognition site of BseRI as that of the first type IIS restriction enzyme and a recognition site of BsgI as that of the second type IIS restriction enzyme and is ligated to the cDNA fragment prepared with Csp6I as type II restriction enzyme, the linker X may be a double-stranded

DNA fragment having the following structure.

5'-···NNNNNNNGAGGAGNGTGCAG···-3' (SEQ ID NO: 13)

3'-···NNNNNNNCTCCTCNCACGTCAT···-5' (SEQ ID NO: 14)

The sequence "5'-GAGGAG-3' "in the linker X is the recognition site of first type IIS restriction enzyme BseRI. The sequence "5'-GTGCAG-3' " at 3' end of the linker X is the recognition site of second type IIS restriction enzyme BsgI. The letter "N" or "n" in base sequence as used herein means any one of bases A, T, C and G.

The term "first type IIS restriction enzyme" as used herein principally means a type IIS restriction enzyme which can recognize the recognition site on linker X and which forms a desired cDNA tag for IEG, or a type I or III restriction enzyme which has the same function as that of the type IIS restriction enzyme.

The first type IIS restriction enzymes used in the invention include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, Bsp423I, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII, BseRI, BspST5I, LweI, PhaI, SfaNI, AarI, Acc36I, BfuAI, BspMI, BveI, Sth132I, SspD5I, AsuHPI, HphI, MboII, NcuI, MnlI, BbsI, BbvII, BbsI, Bbv16II, BpiI, BpuAI, Bsc91I, BspBS31I, BspIS4I, BspTS514I, BstBS32I, BstTS5I, BstV2I, Bme585I, BscAI, Bst19I, BstFZ438I, FauI, SmuI, BciVI, BfuI, and HpyAV.

Among these enzymes, the first type IIS restriction enzymes having a distance of ten or more bases from the recognition site to the farthest end include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, Bsp423I, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII, and BseRI. The first type

IIS restriction enzymes having the distance of 16 or more bases include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI and GsuI.

The term "second type IIS restriction enzyme" as used herein means a type IIS restriction enzyme which can recognize a recognition site on linker X or a recognition site formed in a linking site of linker X-cDNA fragment complex and which cleaves cDNA fragments at an appropriate point, or a type I or III restriction enzyme which has the same function as that of the type IIS restriction enzyme. The linker X-cDNA tag complex is prepared by cleaving with the second type IIS restriction enzyme.

The second type IIS restriction enzymes include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII, BseRI, BspST5I, LweI, PhaI, SfaNI, AarI, Acc36I, BfuAI, BspMI, BveI, Sth132I, SspD5I, AsuHPI, HphI, MboII, NcuI, MnlI, BbsI, BbvII, BbsI, Bbv16II, BpiI, BpuAI, Bsc91I, BspBS31I, BspIS4I, BspTS514I, BstBS32I, BstTS5I, BstV2I, Bme585I, BscAI, Bst19I, BstFZ438I, FauI, SmuI BciVI, BfuI and HpyAV.

Among these enzymes, the second type IIS restriction enzymes having a distance of ten or more bases from the recognition site to the farthest end include

MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII and BseRI. The second type IIS restriction enzymes having the distance of 16 or more bases include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI and GsuI.

Since there is no need to define a sequence of cleavage site formed by the first type IIS restriction enzyme, a combination of the first and second type IIS restriction enzymes is not limited. In addition, although linker X should have the end which may combine with a cDNA fragment formed with the type II restriction enzyme to form the linker X-cDNA fragment complex or the end which can be processed so that it may form the complex, a combination of type II restriction enzyme and the first or the second type IIS restriction enzyme is not limited.

Where the cutting position of positive chain of the second type IIS restriction enzyme is apart from that of complementary chain thereof, linker Y should have a random sequence having the bases, the number of which is the difference between them. The type IIS restriction enzymes falling under the type above include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, RleAI, EciI, TspDTI, TspGWI, Tth111II, BseMII, BseRI, AsuHPI, HphI, MboII, NcuI, MnlI, BciVI, BfuI and HpyAV.

In step (4), the linker X-cDNA fragment complexes are cleaved with the second type IIS restriction enzyme to prepare the linker X-cDNA tag complexes.

For example, where BsgI is used as the second type IIS restriction enzyme, the enzyme recognizes the double-stranded DNA containing the recognition site "5'-GTGCAG-3' " on the linker X-cDNA fragment complex and the complementary sequence and then cleaves the site "5'-GTGCAG-3'(16/14)". That is, BsgI cuts a phosphodiester bond between the bases located at 16 bp and 17 bp 3'- downstream from the base "G" of the 3'-end of recognition site "5'-GTGCAG-3' " and a phosphodiester bond between the bases located at 14 bp and 15 bp 5' upstream from the base "C" of 5'-end of complementary chain "3'-CACGTC-5' " of the recognition site "5'-GTGCAG-3' ". The resultant DNA fragment has the cleavage end having the following structure.

5'-· · · NNNNNNNGAGGAGNGTGCAGTACNNNNNNNNNNNNN -3'

(SEQ ID NO: 15)

3'-· · · NNNNNNNCTCCTCNCACGTCATGNNNNNNNNNNNN-5'

(SEQ ID NO: 16)

In step (5), the linker X-cDNA tag complexes obtained in step (4) by cleaving the linker X-cDNA fragment complexes with the second type IIS restriction enzyme are refined, if necessary. This refinement may be done by, as described in step (3), removing the rest of the cDNA fragments cut away from the

cDNA tags using oligo dT primers. For example, where the oligo dT primer immobilized on latex beads is used in the preparation of the cDNAs, the precipitation of latex beads is exploited, the solution of the cDNAs treated with the restriction enzyme may be centrifuged to precipitate and then remove the cDNA fragments having labeled oligo dT primers. The supernatant from the centrifugation includes the linker X-cDNA tag complexes.

In step (6), the cleavage end of the linker X-cDNA tag complexes formed by the second type IIS restriction enzyme is processed to ligate linker Y having recognition site of the third type IIS restriction enzyme.

In step (7), linker Ys are ligated to the cleavage end of the linker X-cDNA tag complexes to prepare linker X-cDNA tag-linker Y complexes. The term "linker Y" as used herein means a linker having recognition sites of the first or third type IIS restriction enzymes. The recognition sites are preferably located so that the type IIS restriction enzyme may cleave the cDNA tags out without leaving spacer sequences or with leaving spacer sequences at an appropriate point. For example, where the linker X-cDNA fragment complexes are not processed in step (6), the linker Y to be ligated may be the DNA fragment having the following structures. In this example, the linker Y is designed so that two residues A and C may leave as a spacer sequence.

5'- · · · NNNNNNNGGCGGANNNNNNNNGTNN-3' (SEQ ID NO:

17)

3'- · · · NNNNNNNCCGCCTNNNNNNNNNCA -5' (SEQ ID NO: 18)

The linker X-cDNA tag-linker Y complexes prepared in step (7) may have, for example the following sequences.

5'- · · ·

NNNNNNNGAGGAGNGTGCAGTACNNNNNNNNNNNNNNACNNNNNNNNNTCC

GCCNNNNNNN-3' (SEQ ID NO: 19)

3'- · · ·

NNNNNNNCTCCTCNCACGTCATGNNNNNNNNNNNNNNTGNNNNNNNNNAGG

CGGNNNNNN-5' (SEQ ID NO: 20)

The term "third type IIS restriction enzyme" as used herein means a type IIS restriction enzyme, which is used instead of the first type IIS restriction enzyme, having a recognition site on linker Y and being capable of forming a desired cDNA tag, or a type I or III restriction enzyme having the same function as that of the type IIS restriction enzyme.

The third type IIS restriction enzymes include, for example MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, Bsp423I, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII, BseRI, BspST5I, LweI, PhaI,

SfaNI, AarI, Acc36I, BfuAI, BspMI, BveI, Sth132I, SspD5I, AsuHPI, HphI, MboII, NcuI, MnlI, BbsI, BbvII, BbsI, Bbv16II, BpiI, BpuAI, Bsc91I, BspBS31I, BspIS4I, BspTS514I, BstBS32I, BstTS5I, BstV2I, Bme585I, BscAI, Bst19I, BstFZ438I, FauI, SmuI BciVI, BfuI and HpyAV.

Among these enzymes, the third type IIS restriction enzymes having a distance of ten or more bases from the recognition site to the farthest end include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI, GsuI, BsmFI, BspLU11III, BstOZ616I, StsI, BceAI, BstPZ418I, FokI, BcefI, AlwXI, BbvI, Bsp423I, BseKI, BseXI, Bsp423I, Bst12I, Bst71I, BstV1I, RleAI, AceIII, Bbr7I, EciI, TspDTI, TspGWI, Tth111II, HgaI, BseMII and BseRI. The third type IIS restriction enzymes having the distance of 16 or more bases include MmeI, AcuI, Bce83I, BpmI, BpuEI, BsgI, BspKT5I, Eco57I, Eco57MI and GsuI.

A combination of the first and third type IIS restriction enzymes is not limited, even if the both of the enzymes are the same. Where the reaction system is the same or similar to each other among the combinations of the first and the third type IIS restriction enzymes, the both of the enzymes may be reacted simultaneously or the step of removing the enzyme or the salt may be omitted in step (9) described hereinafter.

The step provides with the complex having the structure "5'-[linker X]-[cDNA tag (cDNA tag for IEG)]-[linker Y]-3'".

In step (8), the linker X-cDNA tag-linker Y complexes are amplified. The complexes obtained by step (7) have sequences in linkers X and Y to which primers X and Y may hybridize respectively and may be easily amplified by polymerase chain reaction (PCR). The standard polymerase chain reaction method may be used for the present invention, for example the method described in USP No. 4,683,195. Further the complex may be amplified by cloning the one that is ligated into a vector adaptable to a prokaryote or by another method for the amplification known to those skilled in the art.

Where the PCR is performed using template mixtures comprising a variety of DNAs having different length which are ligated to the linkers for primer annealing at ends thereof, the amplification efficiency varies depending on the length of each template DNA. Generally, as the strands are long, the efficiency of the amplification becomes lower and vice versa. As a result, an occurrence ratio of each amplified fragment in the amplified products thus obtained does not reflect the abundance ratio of correcponding DNA fragment in the mixture of template DNA. In contrast, since the template DNA used in present invention have the same length and is short, the occurrence ratio of each amplified DNA fragment in the resultant amplified products should reflect the abundance ratio of corresponding DNA fragment in the mixture of template DNAs.

Because there is hardly influence due to the difference in amplification efficiency of PCR in theory and therefore the occurrence ratio of each amplified DNA fragment in the resultant amplified products will reflect the ratio of corresponding mRNA expressed in the cells to be examined.

The PCR method may be performed under standard conditions of time and temperature in the present invention. Since the linker X-cDNA tag-linker Y complex used in the invention provides a high efficiency of amplification due to its short and equal in length, the number of annealing/sequence extension cycles may be reduced. In addition, since an efficiency of the PCR method may vary due

to a change in the sequence of linker, appropriate linkers used in the procedure may give a desired efficiency of annealing/sequence extension cycle.

The term "primer X" as used herein means a naturally-occurring or synthesized oligo nucleotide which is complementary to a nucleic acid strand of linker X and may work as an initiation point under conditions that the PCR starts. The primer X should have a length enough to hybridize at the site where a recognition site of the first type IIS restriction enzyme on linker X is retained and to initiate the amplification in the presence of an agent for polymerization. A required length of primer X will be determined due to lots of factors such as temperature, pH and ligase used in the PCR. Likewise, the term "primer Y" as used herein means a naturally-occurring or synthesized oligo nucleotide which is complementary to a nucleic acid strand of linker Y and may work as an initiation point under conditions that the PCR starts. The primer Y should have a length enough to hybridize at the site where recognition sites of the first and third type IIS restriction enzymes on linker Y are retained and to initiate the amplification in the presence of an agent for polymerization.

Those of skill in the art will easily prepare primers for amplification based on the nucleotide sequence of the linkers by taking the first and third type IIS restriction enzymes into consideration without undue experimentation.

In step (9) of Mode 3, the resultant amplified-products are cleaved with

the first and third type IIS restriction enzymes simultaneously or in turn with inserting the additional step of removing the enzyme and salt between the cleaving steps using the first or third enzymes, if necessary, to produce the cDNA tags. The first and third type IIS restriction enzymes may be selected independently. Where the cleavage is performed in turn, either of the enzymes may be used optionally in order of use. For example, where BseRI is used as a first type IIS restriction enzyme, the enzyme recognizes the double-stranded DNA consisting of sequence "5'-GAGGAG-3'" on the linker X and its complementary strand and then cleaves the site "5'-GAGGAG-3'(10/8)". Namely, BseRI cuts a phosphodiester bond between the bases located at 10 bp and 11 bp 3'-downstream from the base "G" of the 3'-end of recognition site "5'-GAGGAG-3'" and a phosphodiester bond between the bases located at 8 bp and 9 bp 5'-upstream from the base "C" of 5'-end of complementary chain "3'-CTCCTC-5'" of the recognition site "5'-GAGGAG-3' ". The resultant DNA fragment of linker X with the cleavage end having the following structure is prepared.

5'- · · · NNNNNNNGAGGAGNGTGCAGTAC-3' (SEQ ID NO: 21)

3'- · · · NNNNNNNCTCCTCNCACGTCA -5' (SEQ ID NO: 22)

Likewise, where EciI is used as a third type IIS restriction enzyme, the enzyme recognizes the double-stranded DNA consisting of sequence "5'-GGCGGA-3'" on the linker Y and its complimentary strand and then cleaves the

site "5'-GGCGGA-3'(11/9)". Namely, EciI cuts a phosphodiester bond between the bases located at 11 bp and 12 bp 3'- downstream from the base "A" of the 3'-end of recognition site "5'-GGCGGA-3'" and a phosphodiester bond between the bases located at 9 bp and 10 bp 5'-upstream from the base "T" of 5'-end of complementary chain "5'-TCCGCC-3'" of the recognition site "5'-GGCGGA-3'". The resultant DNA fragment of linker Y with the cleavage end having the following structure is prepared.

5'- · · · NNNNNNNGGCGGANNNNNNNNNGT -3' (SEQ ID NO: 23)

3'- · · · NNNNNNNCCGCCTNNNNNNNNN -5' (SEQ ID NO: 24)

As a result, the cDNA tag is cut out from the DNA fragments including linkers X and Y.

In short, there is provided the cDNA tag consisting of sequences Nos.25 and 26, comprising a nucleotide chain of thirteen bases immediately adjacent to Csp6I cleavage end (5'-AC-3') of the cDNA fragment derived from the cDNA to be examined, by using Cps6I as a type II restriction enzyme in step (2), linker X comprising a nucleotide chain consisting of base sequences of SEQ ID NOs: 13 and 14 in step (3), BsgI as a second type IIS restriction enzyme in step (4), linker Y comprising a nucleotide chain consisting of base sequences of SEQ ID NOs: 17 and 18 in step (7), and BseRI as a first type IIS restriction enzyme and EciI as a third type IIS restriction enzyme.

5'- NNNNNNNNNNNNAC-3' (SEQ ID NO:25)

3'-TGNNNNNNNNNNNN -5' (SEQ ID NO:26)

Where the method of the present invention is carried out with a cDNA library obtained from mRNAs derived from cells, a library of the cDNA tags is obtained in step (9).

According to the present invention, the cDNAs corresponding to the cDNA tags for identifying expressed genes can be qualitatively or quantitatively detected to analyze a pattern of gene expression by utilizing the resultant library of the cDNA tags.

For example, a selection of target genes can be conducted by providing a detector having spots of a library of cDNA tags corresponding to cDNAs to be detected, contacting each of a sample obtained from a subject and a standard sample, which are labeled with different markers respectively, with the detector and comparing relative signal strength of the different markers. A wide range of known markers such as fluorescent marker and radio isotopic marker may be used in this step.

According to another aspect of the present invention, cDNA tags in a library of the cDNA tags can be detected to analyze a pattern of the gene expression by contacting the library with a detector to which the cDNAs to be detected are immobilized.

The detectors used for the present invention includes a microarray device such as DNA chip and a macroarray device such as dot hybridization. Substrates used for the detector include Nylon membrane, nitrocellulose filter, glass plate and silicon chip. The detector may be, for example a device for detecting target nucleic acids in which the resultant cDNA tags are immobilized on a substrate and then DNA, RNA and/or their fragments to be detected are hybridized thereon.

Samples are preferably labeled in a manner such that mRNAs or cDNAs can be detected. For example, markers in this step include radioisotope, fluorescent compound, bioluminescence compound, chemiluminescence compound, metal chelator or enzyme.

For example, labeled cDNAs to be detected are melted into single strands, if necessary, gradually diluted and then contacted with a solid substrate holding the cDNA tags corresponding to genes to be detected in each grid of silicone chip. Conditions of cell sample can be easily found by comparing the resultant pattern of gene expression with a standard pattern of gene expression. In addition, an expression pattern of unknown gene can be recorded by fixing cDNA tags of the gene. As a result, the gene will be able to be reanalyzed in future, where the gene is identified.

In the present invention, a length of cDNA tag may be adjusted by selecting an appropriate second type IIS restriction enzyme. Although a desired

length of cDNA tag may vary depending on a kind of species to be analyzed, the length of the DNA tag generally ranges from 6 to 25 bp, preferably from 10 to 25 bp and more preferably 10 to 16 bp.

In step (10), the cDNA tag may be isolated, if necessary. The isolation may be conducted by conventional methods used by those skilled in the art such as polyacrylamide gel electrophoresis (PAGE).

In addition, expressed genes may be determined by ligating the cDNA tags to form a concatemer and then sequencing the concatemer. For example, since the cDNA tags obtained in step (9) have 3'- and 5'-cohesive ends which are complementary to each other, they can be ligated each other with T4 ligase. The resultant concatemer of cDNA tags may be analyzed by methods, known to those skilled in the art, for example, cloning into a vector or sequencing with a sequencer.

In the present invention, concatemers generally consist of 3 to 200 of cDNA tags, preferably from 3 to 80 of cDNA tags and more preferably from 16 to 40 of cDNA tags. In this connection, the resultant concatemer may or may not have a spacer sequence between cDNAs tags depending on methods for the preparation of cDNA tags.

The concatemers of the present invention may be cloned by standard methods comprising the steps of integrating the tags into plasmids or phages and

amplifying.

The term "recombinant vector" as used herein refers to a plasmid, virus or other vehicle prepared by inserting or cloning the concatemer of cDNA tags into it. Such a vector includes an origin of replication, a promoter and a specific gene which allows a phenotypic selection of transformed cell. In the present invention, many kinds of known cloning vectors suitable for sequencing may be used. Examples of such vectors include, but are not limited to, pUC18, altered vectors of pUC18 such as pUC118, pUC19, altered vectors of pUC19 such as pUC119, M13mp18RFI, M13mp19RFI, pBR322, pCR3.1, pBAD-TOPO, altered vector of pBAD-TOPO and pBluescript(R)II.

The recombinant vectors are transfected into an appropriate host cell. The term “host cell” as used herein means a cell in which a vector may be amplified and a DNA of the vector may be expressed, or progenies thereof. Since a mutation may occur during their replication, all of the progenies are not always the same as their parent cell.

The present invention may utilize known and stable methods for transferring an exogenous gene by which the gene is continuously retained. For example, where prokaryotic cells such as Escherichia coli are used as a host cell, the cells are collected after the exponential growth phase and treated by know methods such as RbCl method and $CaCl_2$ method to prepare competent cells

having an ability to uptake DNA. The cells may be transformed by electroporation or conventional methods.

According to the present invention, 20 or more cDNA tags, preferably 20 to 100 cDNA tags and more preferably 20 to 30 cDNA tags can be sequenced in an operation by cloning a concatemer of the cDNA tags into a vector and sequencing the concatemer.

Although the preferred embodiments of the present invention have been described herein before, it will be apparent that those skilled in the art may make a variety of changes and modifications without departing from the scope of the present invention. The present invention will be particularly explained on the basis of the following examples, which are not intended to limit a protective scope of the invention. Namely, it should be understood that the present invention will be limited only by the claims attached to the present application.

Examples

(Example 1) Analysis of gene expression of peripheral blood lymphocytes

Peripheral blood mononuclear cells (PBMC) were collected from peripheral blood obtained from normal donor with NycoPrep1.077A (Nyco Med Pharma AS). The resultant peripheral blood lymphocytes were incubated at 37 degrees Celsius for three hours in the presence or absence of 10 ug/ml lipopolysaccharide (LPS), and then total RNA was extracted from the incubated

cells using Isogen (Nippon Gene Co. Ltd.). The total RNA extract obtained was treated at 37 degrees Celsius for 30 minutes with DNaseI (Takara Shuzo Co., LTD) and then refined with RNeasy (QIAGEN). The mRNA was isolated from the total RNA by adsorbing with Oligotex-MAG mRNA refinement kit (Takara Shuzo Co., LTD) and then double-stranded cDNAs were prepared from the mRNA by using cDNA synthesis kit (Takara Shuzo Co., LTD).

The resultant double-stranded cDNAs were cleaved by treating with restriction enzyme Csp6I (MBI Fermentas Inc.) at 37 degrees Celsius for two hours. The cleaved fragments with magnet beads were collected on a wall surface to obtain a fraction including sequences located between a poly A tail of said RNA and the recognition site of Csp6I first appeared in the 5'-upstream direction of the poly A tail by using a magnet. Linker X having a recognition site of a first type IIS restriction enzyme BseRI was ligated to the fraction of the cDNA fragments by one of the three processes described below with T4 DNA ligase.

(1) Process for directly ligating linker X to cleavage end of Csp6I:

Linker X having the following structure was directly ligated to the cohesive end formed by cleaving with Csp6I.

5'-ACCGAGGAGTGTGCAG-3' (SEQ ID NO:27)

3'-TGGCTCCTCACACGTCAT-5' (SEQ ID NO:28)

The linker X-cDNA fragments were cleaved with BsgI (New England

Biolabs) at 37 degrees Celsius for two hours utilizing the recognition site of restriction enzyme BsgI "5'-GTGCAG-3' " in the linker X. The cleaved fragments having no beads, that is, the supernatant was collected. Since the cleavage site of the enzyme was at positon 5'-GTGCAG-3'(16/14), the collected fragments includes 13 base pairs derived from the cDNA following the linker X (except for common three residues "TAC" from Csp6I cleavage site).

Treating the collected fragments for two hours at 16 degrees Celsius with T4 DNA ligase ligated a second linker Y having the structure indicated below.

5'- ACCACTGCGACTCCGCCTGG-3' (SEQ ID NO:29)

3'-NNTGGTGACGCTGAGGCGGACC-5' (SEQ ID NO:30)

By ligating said linker Ys, there was provided a library of complexes consisting of small cDNA fragments "linker X-AC-13 bp derived from cDNA (cDNA tag)-AC-linker Y", that is, a group of DNAs of whole length 52 bp including 13 bp derived from cDNAs interposed between known linkers. This double-stranded fragment consists of the base sequence indicated below and its complementary chain.

5'-

ACCGAGGAGTGTGCAGTACNNNNNNNNNNNNNNACCACTGCGACTCCGCCTG

G-3' (SEQ ID NO:31)

The library of complexes consisting of small cDNA fragments were

amplified by PCR method using Taq DNA polymerase, primer X comprising base sequence "5'-ACCGAGGAGTGTGCAGTA-3'" (SEQ ID NO:32) which hybridize the linker X region and primer Y comprising base sequence "5'-CCAGGCGGAGTCGCAGTGGT-3' " (SEQ ID NO:33) which hybridize the linker Y region. The PCR method was performed by melting at 96 degrees Celsius for 30 seconds, annealing at 50 degrees Celsius for one minute and extending at 72 degrees Celsius for one minutes for 20 cycles and then finally extending at 72 degrees Celsius for two minites.

The obtained PCR products were treated with the type IIS restriction enzymes BseRI (New England Biolabs)and EciI (New England Biolabs). Since the recognition sites of these enzymes are "5'-GAGGAG-3'(10/8)" and "5'-GGCGGA-3'(11/9)" respectively, the DNA fragments having the structure occurred occurred.

5'- NNNNNNNNNNNNNAC-3' (SEQ ID NO:34)

3'-TGNNNNNNNNNNNNN -5' (SEQ ID NO:35)

The treated products were electrophoresed through 12% polyacrylamide gel and isolated from the linker fragments to collect the small fragment DNAs.

After the resultant cDNA tags were ligated each other with T4 ligase to obtain concatemers, they were electrophoresed through 4.5% polyacrylamide gel

to collect the concatemers having 500 bp to 1000 bp in length. The collected concatemers have the structure indicated below. Since 5'-AC-3' adjacent to (N)13 was the sequence derived from Csp6I recognition site of the cDNA, there was provided a library of concatemers of cDNA tags which were completely derived from the cDNAs and which include no spacer sequences artificially added. In the base sequences below," (N)13 "refers to thirteen bases "5'-NNNNNNNNNNNNN-3'" derived from the cDNA.

5'- · · · AC(N)13AC(N)13AC(N)13AC(N)13AC(N)13AC(N)13AC · · · -3'

(SEQ ID NO:36)

3'- · · · TG(N)13TG(N)13TG(N)13TG(N)13TG(N)13TG(N)13TG · · · -5'

(SEQ ID NO:37)

The above concatemer was cloned into plasmid pUC118 and sequenced the base sequence using DNA sequencer (ABI377). As a result, the genes specifically expressed in PBMC or the ones stimulated with LPS were analyzed. It is considered that about 10,000 of tag sequences need to be sequenced in order to approximately identify kinds and estimate appearance ratio of each mRNAs expressed in a cell. Since about 20 cDNA tags could be sequenced in one sequencing operation according to the invention, the kinds of mRNAs and the ratio can be estimated by determining base sequences of about 500 samples.

Tables 1 and 2 show the same genes identified by this method. A homology screening for base sequences of these cDNA tags was carried out by using known database. Table 1 shows the genes which were enhanced by LPS stimulation. Table 2 shows the genes which were suppressed by LPS stimulation.

5

Table 1

Genes whose expressions are enhanced by LPS stimulation

mf ID	mf Base sequence	Name of Gene
11231226	5'–AGGGTCCTTTTGC-3' (SEQ ID No. 38)	hII3.3B Gene for Histon H3.3 (Hs.180877)
65462282	5'-TTGCGTGAAAAGC-3' (SEQ ID No. 39)	Arg-Serpin (plasminogen activator-inhibitor 2, PAI-2) (Hs.75716)
22150632	5'-CCCACTTTCTGCT-3' (SEQ ID No. 40)	Unknown
55149444	5'-TCAGCGAATGAAT-3' (SEQ ID No. 41)	IL-I receptor antagonis, IL-Ira (IL-IRN) Gene, complete codes (IIs.81134)
17350558	5'-CAAGAGTTTGCTC-3' (SEQ ID No. 42)	CC chemokine LARC precursor
58058765	5'-TCTCCTGGAAATA-3' (SEQ ID No. 43)	Cytokine subfamily B (Cys-X-Cys), Member 10(SCYA10) mRNA
27500370	5'-CGGATGCTTCCAC-3' (SEQ ID No. 44)	Interferon Repression factor 1 (IRF-1) mRNA
61929620	5'-TGTAATTGAGCAT-3' (SEQ ID No. 45)	(Putative) Initiation factor (SUI-1) mRNA
49078651	5'-GTGTATGACCTGG-3' (SEQ ID No. 46)	Activation (Act-2) mRNA complete codes (Hs.75703)
24468063	5'-CCTCCCCGGCCTG-3' (SEQ ID No. 47)	JAK Binding protein (SSI-1) mRNA

Table 2

Genes whose expressions are suppressed by LPS stimulation

mf ID	mf Base sequence	Name of Gene
30790136	5'-CTCCCTCACTTCT-3' (SEQ ID No. 48)	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog (FGR) mRNA
32376076	5'-CTGTGAACCAAGT-3' (SEQ ID No. 49)	Liposome protein L3 (RPL3) mRNA
22677064	5'-CCCGGAACGCACT-3' (SEQ ID No. 50)	Major histocompatibility complex class II DM α (HLA-DMA) mRNA
17588982	5'-CAATACGAGTTCC-3' (SEQ ID No. 51)	Actin-related protein 2/3 complex subunit 1B (41Kd)(ARPC1B) mRNA
58325411	5'-TCTGCTTGCGGAG-3' (SEQ ID No. 52)	Homo sapiens zyxin (ZYX) mRNA
22535845	5'-CCCCTTCTGGGCA-3' (SEQ ID No. 53)	G(i) Protein α -subunit (Adenylate cyclase inhibiting GTP- binding protein) (Hs.77269) mRNA
19476075	5'-CAGGCAGTGCGGG-3' (SEQ ID No. 54)	Apoptosis-associated speck-like protein containing a CARD (ASC) mRNA
52161694	5'-TACGTTGTAGCTC-3' (SEQ ID No. 55	Mitochondrial DNA Complete sequence
17076820	5'-CAACAGCAGCCAT-3' (SEQ ID No. 56)	Hematogenesis cell protain-tyrosine kinase (HCK) Gene, Complete sequence Lambda-a2
59268236	5'-TGAGACCTAGAGT-3' (SEQ ID No. 57)	ADP/ATP Translocase mRNA, 3' UTR

5

Numbers designated as mf ID before base sequences in tables 1 and 2 are decimal numbers for computer processing which refer to sequences of 13 bases. Namely, the mf ID is a decimal number generated by substituting 0 for a, 1 for c, 2 for g and 3 for t to make a quaternary digit, converting the number to a decimal number and adding one. Base sequences can be processed as a number regardless of their length. For example, when base sequences consisting of 13 bases are processed, the sequences can be identified by using the numbers in a manner indicated below.

5'-aaaaaaaaaaaaa-3' 00000001 (or simply referred to as 1)

5'-aaaaaaaaaaaac-3' 00000002 (or simply referred to as 2)

5'-aaaaaaaaaaaag-3' 00000003 (or simply referred to as 3)

5'-aaaaaaaaaaaat-3' 00000004 (or simply referred to as 4)

5'-aaaaaaaaaaaca-3' 00000005 (or simply referred to as 5)

.....

.....

5'-ttttttttttgt-3' 67108860

5'-tttttttttttta-3' 67108861

5'-ttttttttttttc-3' 67108862

5'-tttttttttttg-3' 67108863

5'-ttttttttttttt-3' 67108864

By using such IDs, any sequences consisting of 13 bases may be designated one ID number of eight digits. These figures are referred to as mini fragment ID (mf ID).

Example 2

The library of the cDNA tags prepared in example 1 is detected with the detector described below to analize the gene expression.

A DNA chip is produced by synthesizing oligo DNAs comprising sequences corresponding to mf base seqences designated as mf IDs 65462282, 55149444, 17350558, 58058765, 27500370 and 49078651 of the genes listed in Table 1 whose expression was activated by LPS stimulation, and spotting on a slide glass with the oligo DNAs using a conventional method.

In order to prepare the probe solutions, mRNAs derived from the peripheral blood mononuclear cells (PBMC) obtained by LPS stimulation in example 1 which are used as a template were labeled with a fluorescent marker, fluorescent compound Cy3-dUTP (*1) (Amersham Pharmacia), and other mRNAs derived from PBMC not stimulated with LPS which were used as a template were labeled with a fluorescent marker, fluorescent compound Cy5-dUTP (*5) (Amersham Pharmacia).

The probe solutions are mixed together to make 6×SET [0.9M NaCl, 10 μg/ml Yeast tRNA, 0.1%SDS, 120mM Tris-HCl(pH7.8)] solution and then keep in contact with said oligo DNA chip at 45 degrees Celsius overnight to perform hybridization.

After the DNA chip is washed with a washing liquid [6×SSC, 0.1%SDS]at 52 degrees Celsius, the fuluorescent markers on the chip are scanned with a scanner to obtain the fluorescence intensity data and then the data is analyzed. The scatter plot of signal intensity of Cy3 and Cy5 at each of the spots demonstrates that the fluorescent light radiated by the probes derived from the mRNAs from the PBMC stimulated with LPS is more than twice stronger than that from the PBMC not stimulated with LPS at all of the spots.

*1) CAS RN Cy3 CAS RN 146368-16-3

CN 3H-Indolium, 2-[3-[1-[6-[(2,5-dioxo-1-pyrrolidinyl)oxy]-6-oxohexyl]-1,3-dihydro-3,3-dimethyl-5-sulfo-2H-indol-2-ylidene]-1-propenyl]-1-ethyl-3,3-dimethyl-5-sulfo-, inner salt (9CI) (CA INDEX NAME)

*2) CAS RN Cy5 CAS RN 146368-14-1

CN 3H-Indolium, 2-[5-[1-[6-[(2,5-dioxo-1-pyrrolidinyl)oxy] -6- oxohexyl]-1,3-dihydro-3,3-dimethyl-5-sulfo-2H-indol-2-ylidene]-1,3-pentadienyl]-1-ethyl-3,3-dimethyl-5-sulfo-, inner salt (9CI) (CA INDEX NAME)

Example 3

By using the cDNA tags optionally selected from the library of the cDNA tags obtained in example 1, gene expression differences between a pair of samples can be analyzed.

cDNAs prepared with a reverse transcriptase by using, mRNAs derived from peripheral blood mononuclear cells (PBMC) stimulated with LPS and mRNAs derived from PBMC not stimulated with LPS as a template, are spotted on a nylon membrane and then the cDNAs on the membrane are incubated at 80 degrees Celsius for two hours.

An oligo DNA comprising the sequence of mfID65462282 selected from the genes whose expression are induced by LPS stimulation as shown in Table 1 is synthesized and then labeled with [gamma-^{32}P] ATP (Amersham Pharmacia) by T4 polynucleotide kinase to obtain a probe solution including the probe labeled with ^{32}P (radioisotope).

This probe solution is used to perform a hybridization with said nylon membrane in 6 x SET overnight at 45 degrees Celsius. After the nylon membrane is washed with washing solution [6 x SSC, 0.1% SDS] at 52 degrees Celsius, and then performed autoradiography. The signals on X-ray film of the cDNA derived from the mRNA of LPS stimulated PBMC is twice stronger than those of LPS

non-stimulated PBMC.

Example 4

A library of cDNA tags for identifying expressed genes was prepared in the same manner used in example 1 except that AfaI was used instead of Csp6I as a type II restriction enzyme and BseRI was used instead of EciI as a third type IIS restriction enzyme.

(1) The cDNAs in the sample were directly ligated to linker Xes having the structure illustrated below at the blunt end of the cDNA which was formed by cleaving with type II restriction enzyme AfaI.

5'- ACCGAGGAGTGTGCAGT-3' (SEQ ID NO:58)

3'- TGGCTCCTCACACGTCA-5' (SEQ ID NO:59)

(2) Base "A" was ligated to the blunt end of the cDNA in the sample formed by cleaving with type II restriction enzyme AfaI in the presence of dATP in the following manner.

5'- AC · · · · -3

3'- ATG · · · · -5'

In the next step, the cohesive end above was ligated to linker X having the following structure.

5'- ACCGAGGAGTGTGCAGT-3' (SEQ ID NO:60)

3'- TGGCTCCTCACACGTC-5' (SEQ ID NO:61)

(3) Base "T" was deleted from the blunt end of the cDNA in the sample formed by cleaving with type II restriction enzyme AfaI in the presence of dGTP in the following manner.

5'- AC · · · · -3'

3'- G · · · · -5'

In the following step, the cohesive end above was ligated to linker X having the following structure.

5'- ACCGAGGAGTGTGCAGT -3' (SEQ ID NO:62)

3'- TGGCTCCTCACACGTCAT-5' (SEQ ID NO:63)

The linker X-cDNA fragments were cleaved with BsgI by utilizing the

recognition site of restriction enzyme BsgI "5'-GTGCAG-3' " in the linker X. The mixture of cleaved products were centrifuged and the supernatant was collected. Since the cleavage site of the enzyme is at positon 5'-GTGCAG-3'(16/14), the collected fragments includes tags of 13 base pairs derived from the cDNA following the linker X.

Subsequently, linker Ys having the structure illustrated below were ligated. A desired library of the cDNA tags were obtained by treating in the same manner as described in example 1 and digesting with the first and third type IIS restriction enzymes.

5' - ACCACTGCGACTCCTCTGG-3' (SEQ ID NO:64)

3' - NNTGGTGACGCTGAGGAGACC-5' (SEQ ID NO:65)

Example 5

The gene expression may be analyzed by detecting the library of the cDNA tags obtained in example 4 with a detector described below.

A DNA chip is produced by synthesizing oligo DNAs corresponding to mf IDs 65462282, 55149444, 17350558, 58058765, 27500370 and 49078651 of the genes listed in Table 1 whose expression is activated by LPS stimulation, and spotting on a slide glass with the oligo DNAs using a conventional method.

A probe solution is prepared by labeling mRNAs derived from peripheral blood mononuclear cells (PBMC) obtained by LPS stimulation in example 1 which are used as a template with a fluorescent marker, fluorescent compound Cy3-dUTP (Amersham Pharmacia)(*1: See example 2), and labeling other mRNAs derived from PBMC not stimulated with LPS which are used as a template with a fluorescent marker, fluorescent compound Cy5-dUTP (Amersham Pharmacia)(*2: see example 2).

The probe solutions are mixed together to make 6×SET [0.9M NaCl, 10 μg/ml Yeast tRNA, 0.1%SDS, 120mM Tris-HCl(pH7.8)] solution and then kept in contact with said oligo DNA chip at 45 degrees Celsius overnight to perform hybridization.

After the DNA chip is washed with a washing solution [6×SSC, 0.1 %SDS]at 52 degrees Celsius, the fluorescent markers on the chip are scanned with a scanner to obtain the fluorescence intensity data and then the data was analyzed. The scatter plot of signal intensity of Cy3 and Cy5 at each of the spots demonstrates that the fluorescent light radiated by the probes derived from the mRNAs from the PBMC stimulated with LPS is more than twice stronger than that from the PBMC not stimulated with LPS at all of the spots.

Effect of the Invention

According to the present invention, cDNAs to be tested or genes specifically expressed in cells to be tested can be accurately detected with a high reproducibility to analyze. A method of the present invention can indicate differences of gene expression between optional two kinds of cells to clarify differences in their functions and morphologies. The method is therefore applicable to analysis of huge aspect of biological phenomena under a physiological condition or a diseased state. The present inventors developed a method for the preparation of cDNA tags for identifying expressed genes and a method for the analysis of gene expression, and applied as PCT application No. PCT/JP02/02338 for the invention before filing this application. The present invention provides the method for the preparation of the cDNA tags and the method for the analysis of gene expression which are improved in that a wider selection of the restriction enzymes used may be allowed than those of the prior invention and the preparation of the linkers used may be prepared easier than those used in the prior invention.